Semantic Web Tutorial and Infusion Roadmap (Draft for NASA/DSWG/TIWG, version V0.5, 20070103)

Peter Fox (NCAR)

ESIP Federation Winter Meeting

Jan 3-5, 2007 Portland, OR

Some material borrowed from a presentation by Ivan Herman (W3C)

http://www.w3.org/2006/Talks/1016-Beijing-IH/

WARNING - this talk contains acronyms

Definitions

- Semantic Web
  - An extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation, www.semanticweb.org
    - Primer: http://www.ics.forth.gr/isl/swprimer/
- Semantic Grid
  - Semantic services to use the resources of many computers connected by a network to solve large scale computational problems
- Ontology (n.d.).
  - An explicit formal specification of how to represent the objects, concepts and other entities that are assumed to exist in some area of interest and the relationships that hold among them.
  - The Free On-line Dictionary of Computing. http://dictionary.reference.com/browse/ontology
- DL - Description Logic
- FOL - First Order Logic
- RDF - Resource Description Framework
- OWL - Ontology Web Language (Lite, DL, Full)

Application Areas for SW

- Smart search
- Annotation (even simple forms)
- Geospatial
- Data integration
- Web content mining with natural language parsing
- User interface development (portals)
- Semantic desktop
- Wikis - OntoWiki, SemanticMediaWiki
- Sensor Web
- Software engineering
- Implementing logic (rules), e.g. in workflows
- Web services
- Explanation
- Verification …. and the list goes on

Semantic Web Myths

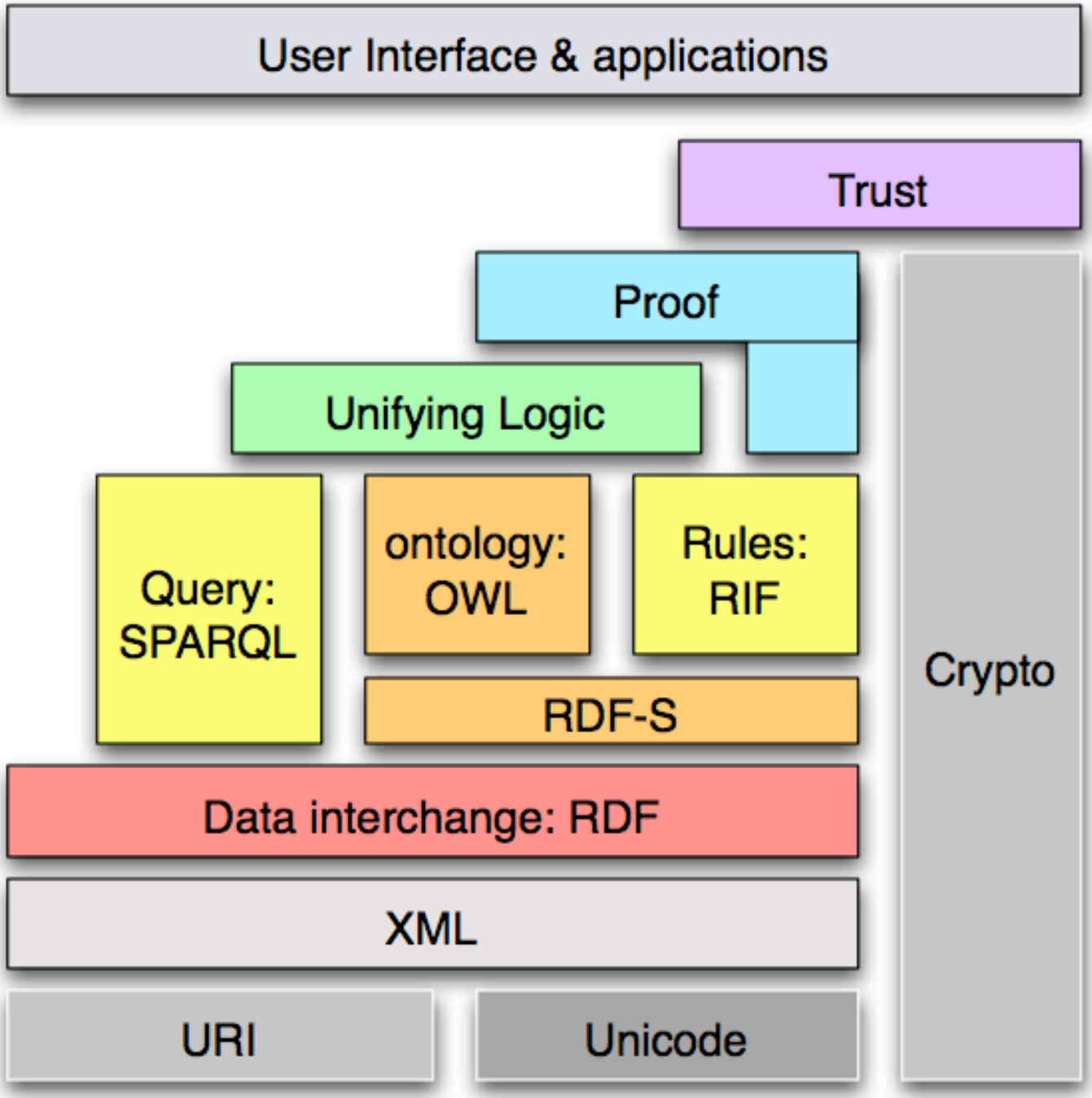
- ‘the Semantic Web is a reincarnation of Artificial Intelligence on the Web’
- ‘it relies on giant, centrally controlled ontologies for "meaning" (as opposed to a democratic, bottom-up control of terms)’
- ‘one has to add metadata to all Web pages, convert all relational databases, and XML data to use the Semantic Web’
- ‘it is just an ugly application of XML’
- ‘one has to learn formal logic, knowledge representation techniques, description logic, etc, to use it’
- ‘it is, essentially, an academic project, of no interest for industry’

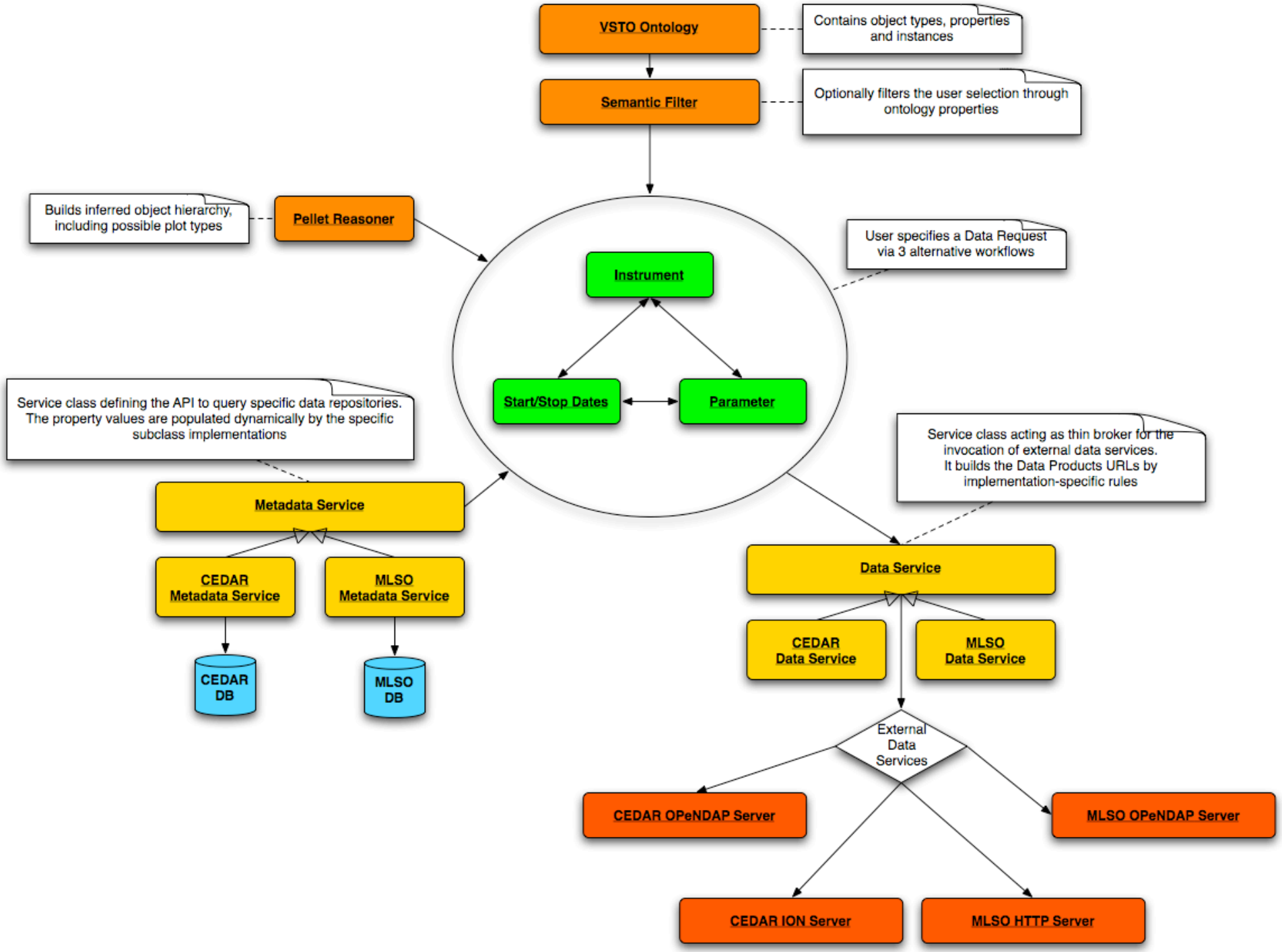
SW != ontologies on the web (!)

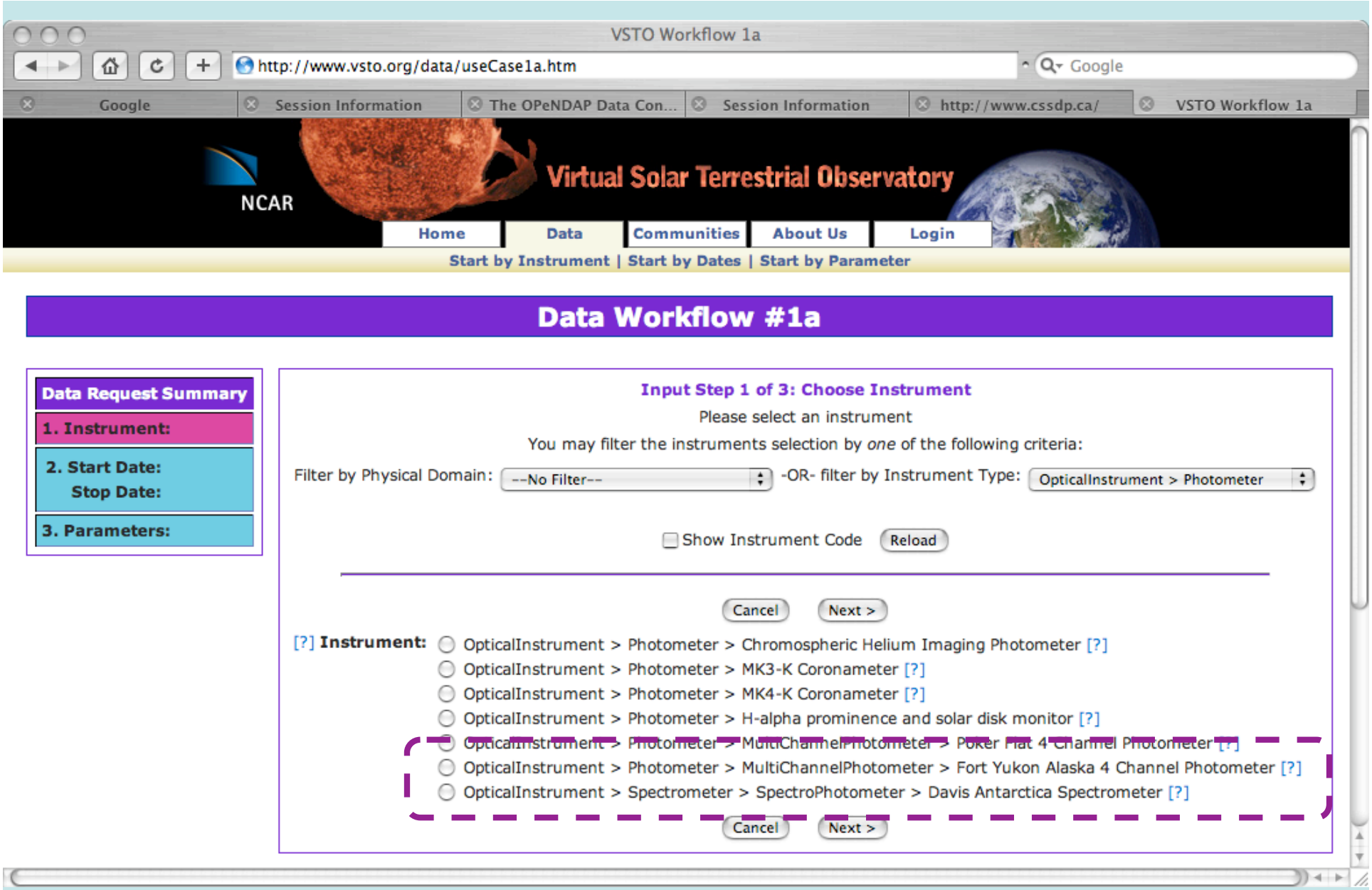
- Ontologies are important, but use them only when necessary
- You can be a perfectly decent citizen of the Semantic Web if you do not use Ontologies, not even RDFS!
- The Semantic Web is about integrating data on the Web; ontologies (and/or rules) are tools to achieve that when necessary
- SW ontologies != some big (central) ontology
  - The ethos of the Semantic Web is on sharing, ie, sharing possibly small ontologies
  - A huge, central ontology would be unmanageable OWL includes statements for versioning, for equivalence and disjointness of terms
- The practice:
  - SW applications using ontologies always mix large number of ontologies and vocabularies (FOAF, DC, and others)
  - the real advantage comes from this mix: that is also how new relationships may be discovered
- http://www.metamodel.com/article.php?story=20030115211223271

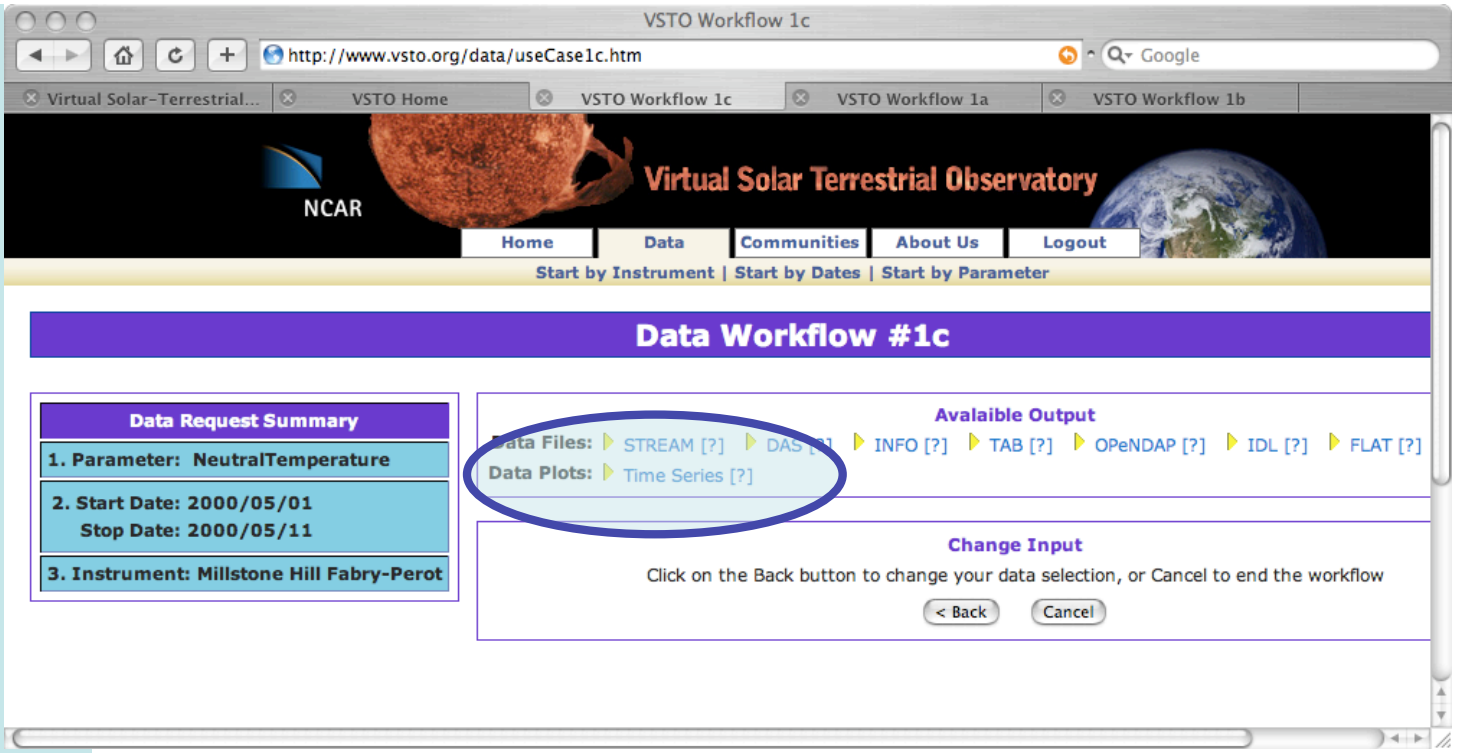
Semantic Web Basics

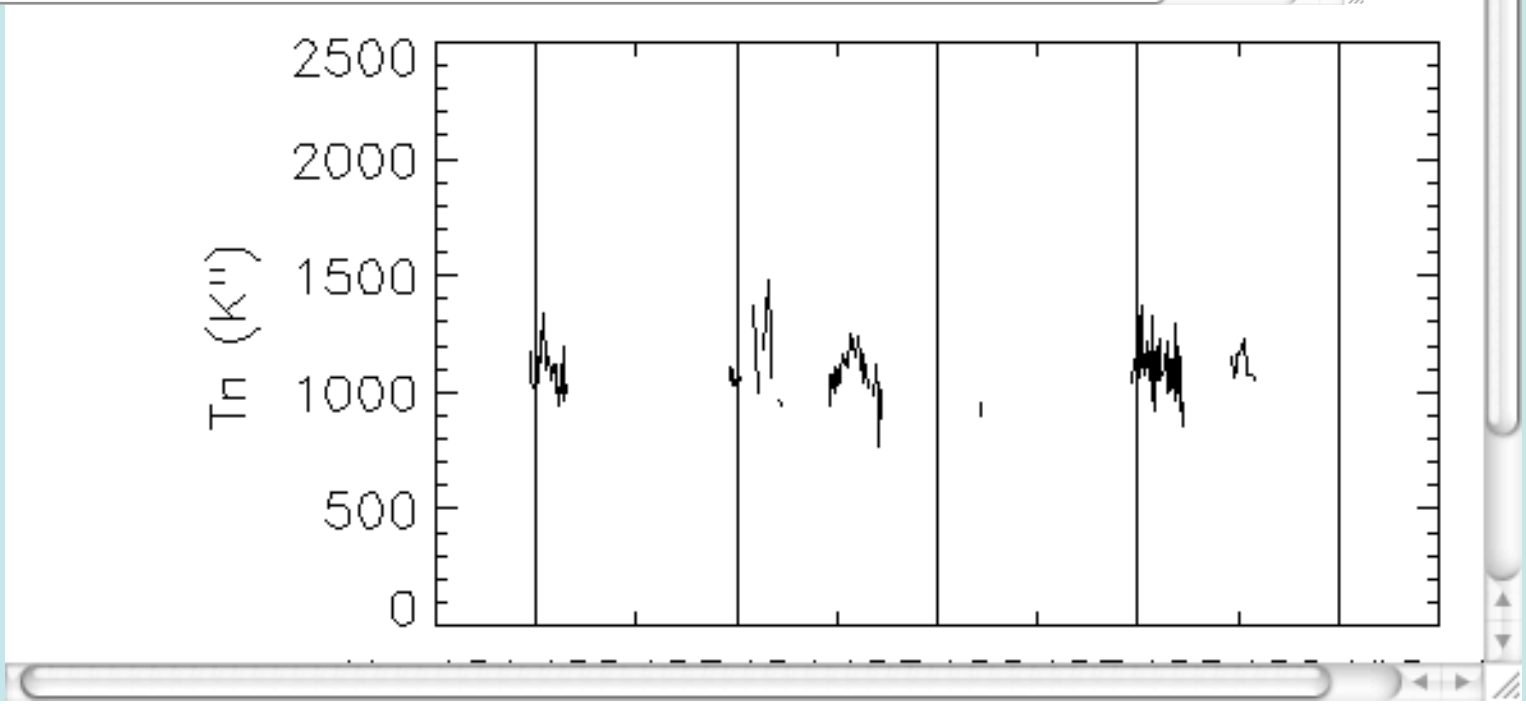
- Lots of tools are available. Are listed on W3C's wiki: RDF programming environment for 14+ languages, including C, C++, Python, Java, Javascript, Ruby, PHP,...(no Cobol or Ada yet ;-( )
- 13+ Triple Stores, ie, database systems to store (sometimes huge!) datasets
- Some of the tools are Open Source, some are not; some are very mature, some are not: it is the usual picture of software tools, nothing special any more!
- Anybody can start developing RDF-based applications today
- OWL 1.0 - Ontology Web Language is a W3C standard, OWL 1.1 on the way


User Interface & applications
Trust
Proof
Unifying Logic
Query: SPARQL
ontology: OWL
Rules: RIF
RDF-S
Crypto
Data interchange: RDF
XML
URI
Unicode


VSTO Ontology
Contains object types, properties and instances
Semantic Filter
Optionally filters the user selection through ontology properties
Builds inferred object hierarchy, including possible plot types
Pellet Reasoner
User specifies a Data Request via 3 alternative workflows
Instrument
Start/Stop Dates
Parameter
Service class defining the API to query specific data repositories. The property values are populated dynamically by the specific subclass implementations
Service class acting as thin broker for the invocation of external data services. It builds the Data Products URLs by implementation-specific rules
Metadata Service
CEDAR Metadata Service
MLSO Metadata Service
CEDAR DB
MLSO DB
Data Service
CEDAR Data Service
MLSO Data Service
External Data Services
CEDAR OPeNDAP Server
MLSO OPeNDAP Server
CEDAR ION Server
MLSO HTTP Server


VSTO Workflow 1a
http://www.vsto.org/data/useCase1a.htm
Google
Google
Session Information
The OPeNDAP Data Con...
Session Information
http://www.cssdp.ca/
VSTO Workflow 1a
NCAR
Virtual Solar Terrestrial Observatory
Home
Data
Communities
About Us
Login
Start by Instrument | Start by Dates | Start by Parameter
Data Workflow #1a
Data Request Summary
1. Instrument:
2. Start Date:
Stop Date:
3. Parameters:
Input Step 1 of 3: Choose Instrument
Please select an instrument
You may filter the instruments selection by one of the following criteria:
Filter by Physical Domain: --No Filter--
-OR- filter by Instrument Type: OpticalInstrument > Photometer
Show Instrument Code
Reload
Cancel
Next >
[?] Instrument:
OpticalInstrument > Photometer > Chromospheric Helium Imaging Photometer [?]
OpticalInstrument > Photometer > MK3-K Coronameter [?]
OpticalInstrument > Photometer > MK4-K Coronameter [?]
OpticalInstrument > Photometer > H-alpha prominence and solar disk monitor [?]
OpticalInstrument > Photometer > MultiChannelPhotometer > Poker Flat 4 Channel Photometer [?]
OpticalInstrument > Photometer > MultiChannelPhotometer > Fort Yukon Alaska 4 Channel Photometer [?]
OpticalInstrument > Spectrometer > SpectroPhotometer > Davis Antarctica Spectrometer [?]
Cancel
Next >


VSTO Workflow 1c
http://www.vsto.org/data/useCase1c.htm
Google
Virtual Solar-Terrestrial...
VSTO Home
VSTO Workflow 1c
VSTO Workflow 1a
VSTO Workflow 1b
NCAR
Virtual Solar Terrestrial Observatory
Home
Data
Communities
About Us
Logout
Start by Instrument | Start by Dates | Start by Parameter
Data Workflow #1c
Data Request Summary
1. Parameter: NeutralTemperature
2. Start Date: 2000/05/01
Stop Date: 2000/05/11
3. Instrument: Millstone Hill Fabry-Perot
Avalaible Output
Data Files: STREAM [?] DAS [?] INFO [?] TAB [?] OPeNDAP [?] IDL [?] FLAT [?]
Data Plots: Time Series [?]
Change Input
Click on the Back button to change your data selection, or Cancel to end the workflow
< Back
Cancel


2500
2000
1500
1000
500
0
Tn (K)

Virtual Solar-Terrestrial... | VSTO Home | VSTO Workflow 1c | VSTO Workflow 1a | VSTO Workflow 1b

NCAR

Virtual Solar Terrestrial Observatory

Home | Data | Communities | About Us | Login

Start by Instrument | Start by Dates | Start by Parameter

Data Workflow #1c

Data Request Summary

1. Parameter:
2. Start Date:
   Stop Date:
3. Instrument:

Input Step 1 of 3: Choose Parameter

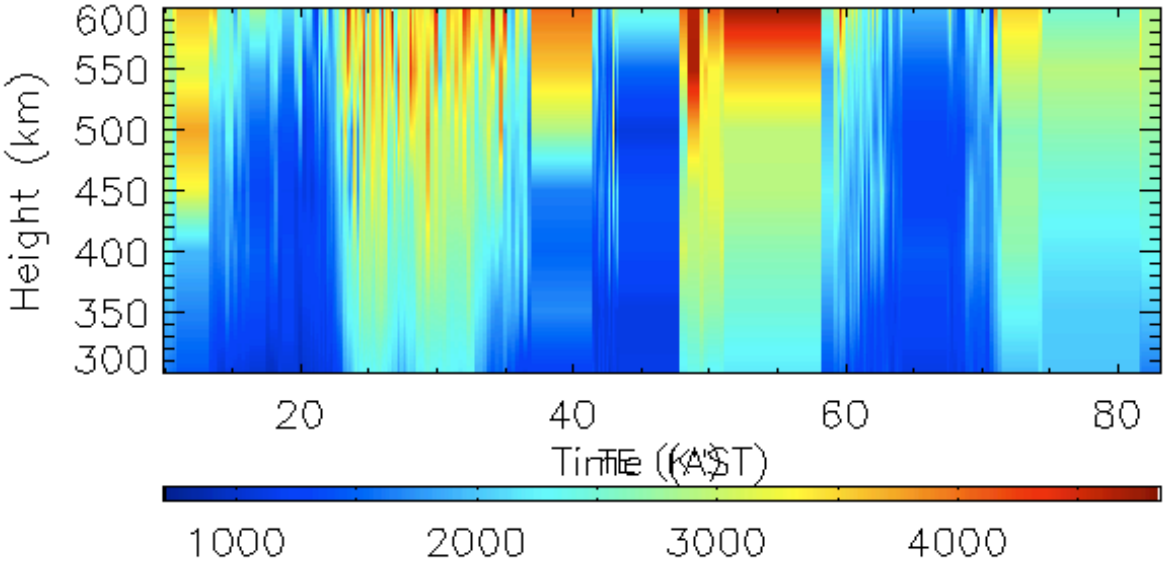
Please select one parameter from the list

You may filter the parameters selection by *one* of the following criteria:

Filter by Physical Domain: --No Filter-- Filter by Parameter Type: Temperature

[?] Parameter: Temperature > ElectronTemperature

Cancel Next >



E.g.’s of what is encoded

- Large datasets are accumulating. E.g.:
  - IngentaConnect bibliographic metadata storage: over 200 million triplets
  - RDF version of Wikipedia: more than 47 million triplets
  - Tracking the US Congress: data stored in RDF (around 25 million triplets)RDFS/OWL
  - Representation of Wordnet: also downloadable as 150MB of RDF/XML
  - Departement/canton/commune structure of France published by the French Statistical Institute
- Large ontologies are being developed (converted from other formats or defined in OWL)
  - eClassOwl: eBusiness ontology for products and services, 75,000 classes and 5,500 properties
  - The Gene Ontology: to describe gene and gene product attributes in any organism
  - UniProt: protein sequence and annotation terminology and data
- However, they do not have to be ‘large’

What about Earth Science?

- SWEET (Semantic Web for Earth and Environmental Terminology)
  - http://sweet.jpl.nasa.gov
  - based on GCMD terms
  - modular using faceted and integrative concepts
- VSTO (Virtual Solar-Terrestrial Observatory)
  - http://vsto.hao.ucar.edu
  - captures observational data (from instruments)
  - modular using domains
- MMI
  - http://marinemetadata.org
  - captures aspects of marine data, ocean observing systems
  - partly modular, mostly by developed project
- GeoSciML
  - http://www.opengis.net/GeoSciML/
  - is a GML (Geography ML) application language for Geoscience
  - modular, in ‘packages’

Best practices

- Modular ontologies…. Driven by use cases
- Rapid prototyping using common tools
- Go “Lite” as much as possible, then DL and *only* if you have to Full - balancing expressibility vs. implementability
- RDF (the model) is not RDF/XML (the encoding)
- There are also a number ‘core vocabularies’ (not necessarily OWL based)
  - SKOS Core: about knowledge systems
  - Dublin Core: about information resources, digital libraries, with extensions for rights, permissions, digital right management
  - FOAF: about people and their organizations
  - DOAP: on the descriptions of software projects
  - MusicBrainz: on the description of CDs, music tracks, ...
  - SIOC: Semantically-Interlinked Online Communities...
- One should never forget: ontologies/ vocabularies must be shared and reused! swoogle.umbc.edu, www.planetont.org

Best Practices - ctd.

- The W3C SW Best Practices and Deployment Working Group (http://www.w3.org/2001/sw/BestPractices/) has developed some documents:
  - Best Practice Recipes for Publishing RDF Vocabularies
  - Defining N-ary relations
  - Representing Classes As Property Values
  - Representing "value partitions" and "value sets"
  - XML Schema Datatypes in RDF and OWL
- The work is continuing in the (new) SW Deployment Working Group (http://www.w3.org/2006/07/SWD/)

Query

- Querying RDF graphs becomes essential
  - SPARQL is almost here query language based on graph patterns
  - There is also a protocol layer to use SPARQL over, eg, HTTP
  - There are a number of implementations already
  - There are also SPARQL 'endpoints' on the Web: send a query and a reference to data over HTTP GET, receive the result in XML or JSON
  - applications may not need any direct RDF programming any more, just a SPARQL endpoint
  - http://www.sparql.org/ and http://www.w3.org/TR/rdf-sparql-query/
- Not a W3C recommendation yet (mid-2007?)!
- Some features are missing
  - query on list/sequence/set membership
  - control and/or description on the entailment regimes of the triple store (RDFS? OWL-DL? OWL-Lite? ...)
  - modify the triple store
- Other options XQUERY, SeRQL, OWL-QL, RDFQuery

Reasoners (aka Inference engines)

- Pellet
- Racer (and Racer Pro)
- Medius KBS
- FACT++
- fuzzyDL
- KAON2
- MSPASS
- QuOnto
- Jess (for Rules)

‘Collecting’ the ‘data’

- Part of the (meta)data information is present in tools ... but thrown away at output e.g., a business chart can be generated by a tool: it ‘knows’ the structure, the classification, etc. of the chart,but, usually, this information is lost storing it in web data would be easy!
- SW-aware tools are around (even if you do not know it...), though more would be good:
  - Photoshop CS stores metadata in RDF in, say, jpg files (using XMP)
  - RSS 1.0 feeds are generated by (almost) all blogging systems (a huge amount of RDF data!)
- Scraping - different tools, services, etc, come around every day:
  - get RDF data associated with images, for example: service to get RDF from flickr images
  - service to get RDF from XMP
  - XSLT scripts to retrieve microformat data from XHTML files
  - scripts to convert spreadsheets to RDF
- SQL - A huge amount of data in Relational Databases
  - Although tools exist, it is not feasible to convert that data into RDF
  - Instead: SQL ⇋ RDF ‘bridges’ are being developed: a query to RDF data is transformed into SQL on-the-fly
  - the modalities are governed by small, local ontologies or rules

Future Collecting

- RDFa (formerly known as RDF/A) extends XHTML by:
  - extending the ***link*** and ***meta*** to include child elements
  - add metadata to any elements (a bit like the ***class*** in microformats, but via dedicated properties)
- It is very similar to microformats, but with more rigor:
  - it is a general framework (instead of an ✗agreement⊤ on the meaning of, say, a ***class*** attribute value)
  - terminologies can be mixed more easily
- The W3C Working Group on SW Deployment has this on its charter
- May be considered as an alternative serialization of (part of) RDF; may be bound to GRDDL in practice

Rules (aka ‘Logic’)

- OWL-DL and OWL-Lite are based on Description Logic
- There are things that DL cannot express (though there are things that are difficult to express with rules and easy in DL...)
  - A well known examples is Horn rules (eg, the ‘uncle’ relationship): (P1 ∧ P2 ∧ ...) → C
  - e.g.: for any X, Y and Z: if Y is a parent of X, and Z is a brother of Y then Z is the uncle of X
- Several attempts already to combine Semantic Web with Rules (Metalog, RuleML, SWRL, WRL, cwm, ...)

Services

- Ontologies of services, provides:
  - *What does the service provide for prospective clients?* The answer to this question is given in the "profile," which is used to advertise the service. To capture this perspective, each instance of the class Service presents a ServiceProfile.
  - *How is it used?* The answer to this question is given in the "process model." This perspective is captured by the ServiceModel class. Instances of the class Service use the property describedBy to refer to the service's ServiceModel.
  - *How does one interact with it?* The answer to this question is given in the "grounding." A grounding provides the needed details about transport protocols. Instances of the class Service have a supports property referring to a ServiceGrounding.

Services, not standard…

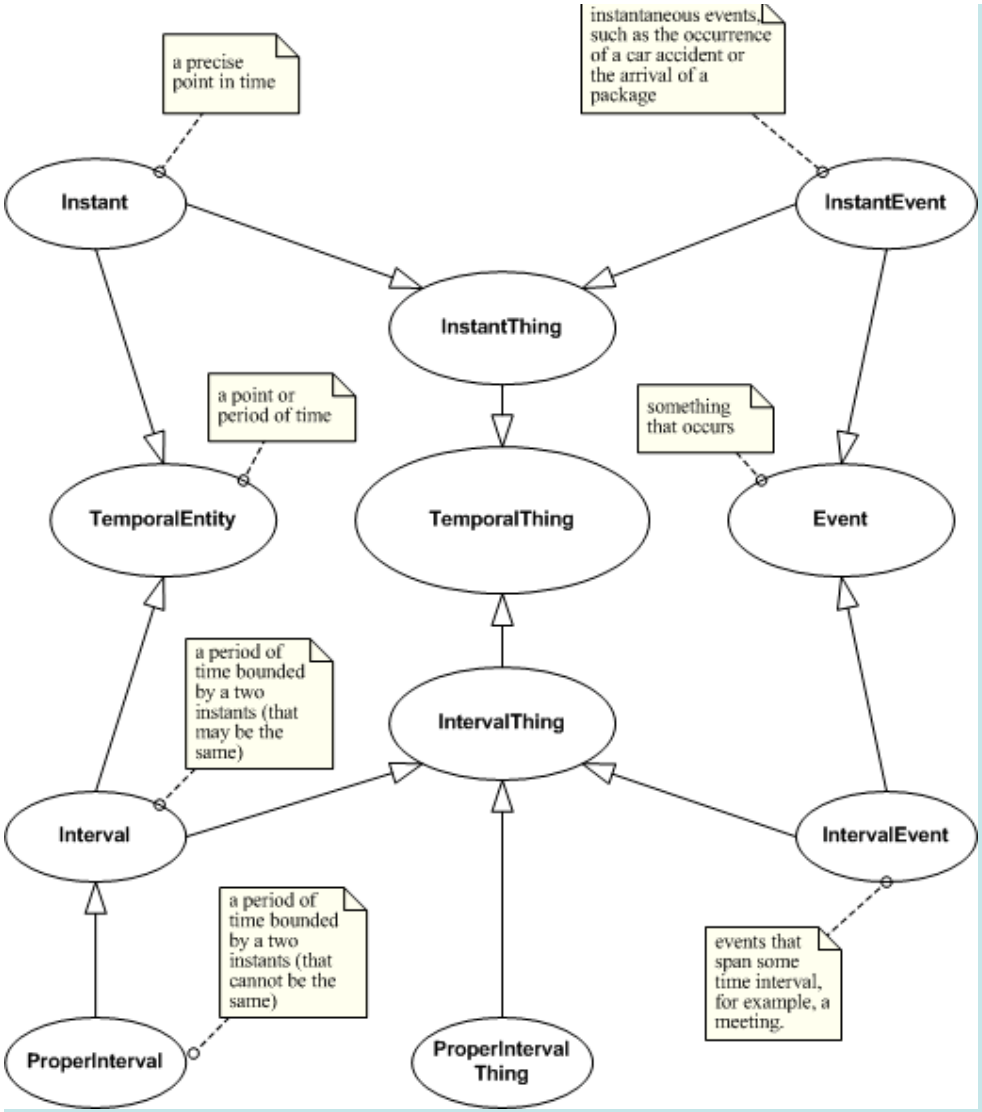
- Submissions to W3C
  - OWL-S - http://www.w3.org/Submission/OWL-S
  - SWSO/F/L - Semantic Web Services Ontology/Framework/Language - http://www.w3.org/Submission/SWSF/
  - WSMO/X/L - Web Services Modeling Ontology/Exection/Language - http://www.w3.org/Submission/WSMX/ www.wsmo.org, www.wsmx.org

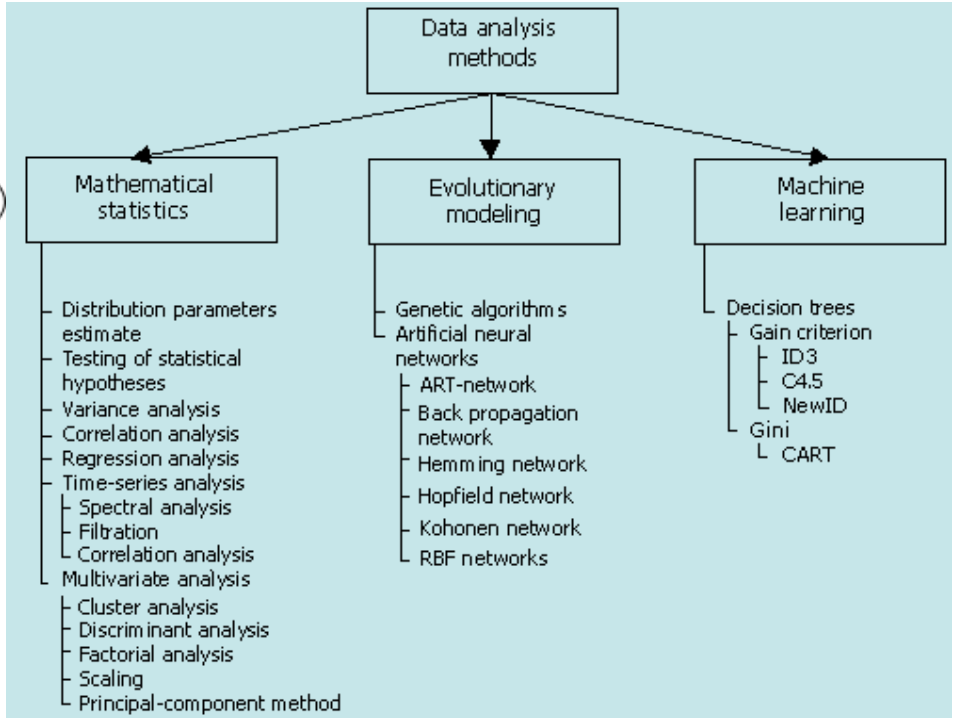
Explanation, Proof (path to Trust)

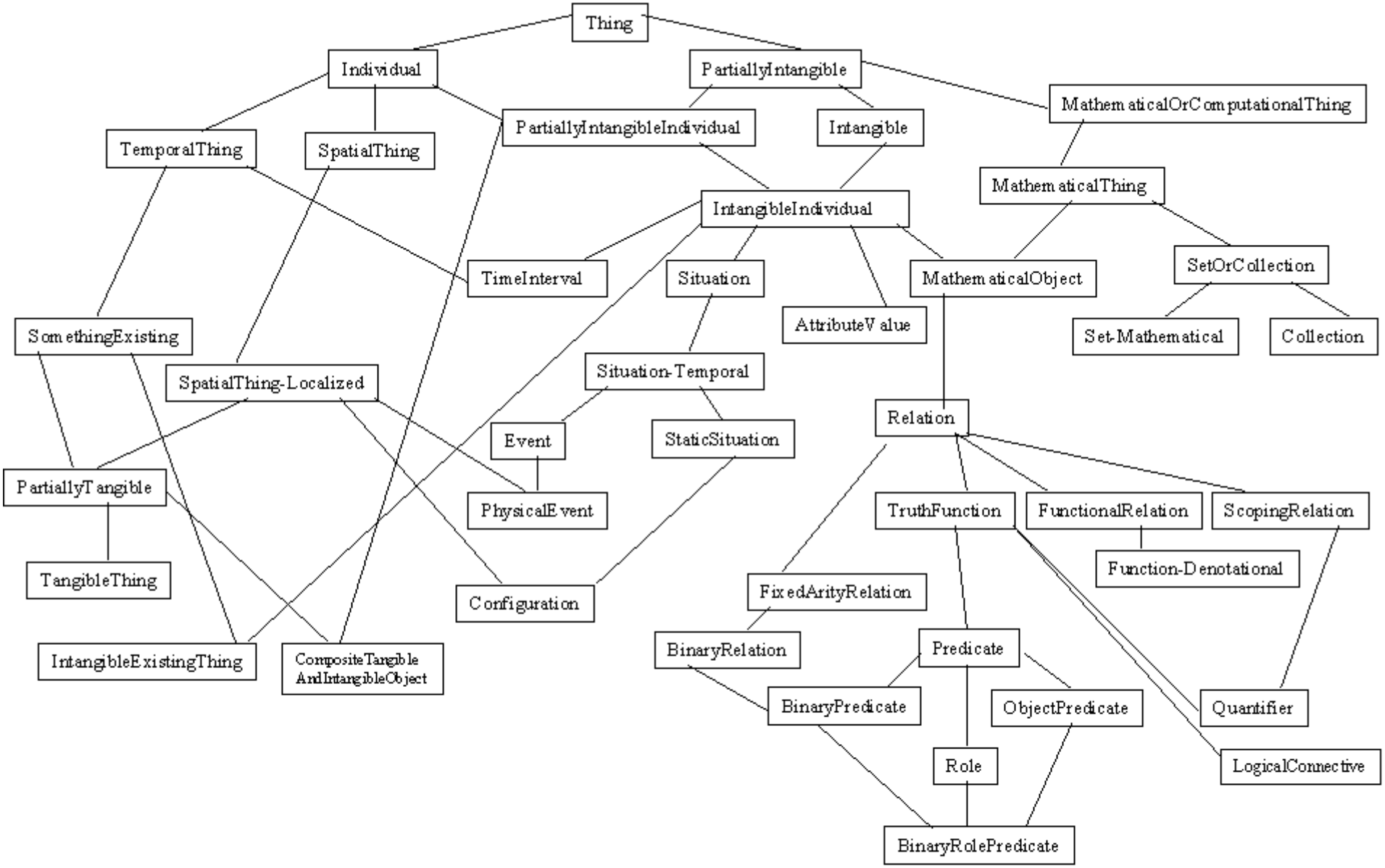
- Proof markup language (PML)
  - an interlingua representation for justifications of results produced by Semantic Web services
- Not W3C, but no competition
- Implemented in InferenceWeb (http://iw.stanford.edu)
- CWM and N3 and theorem provers - not yet adapted to OWL-based languages

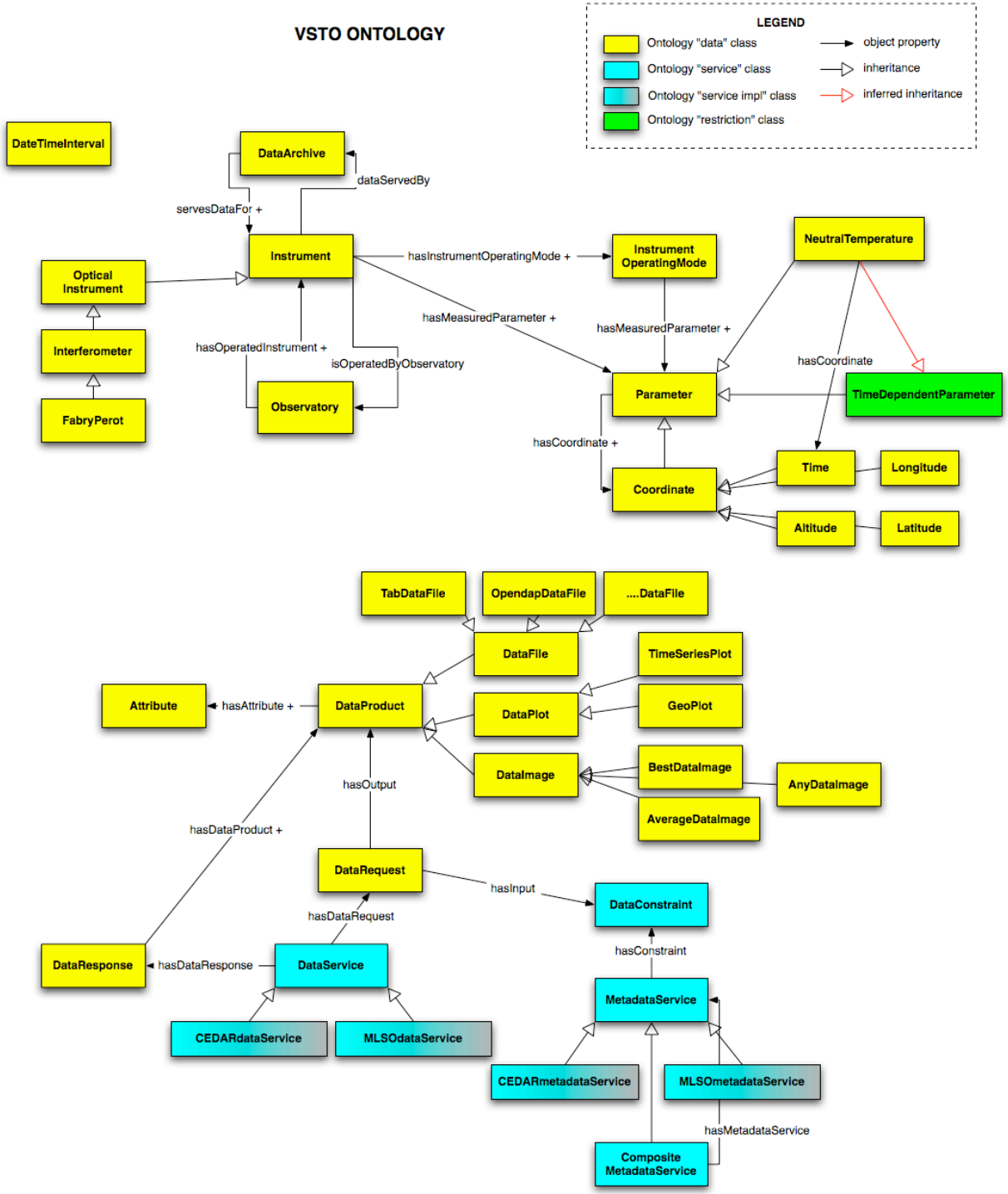
Creating Ontologies - UML, CMAP

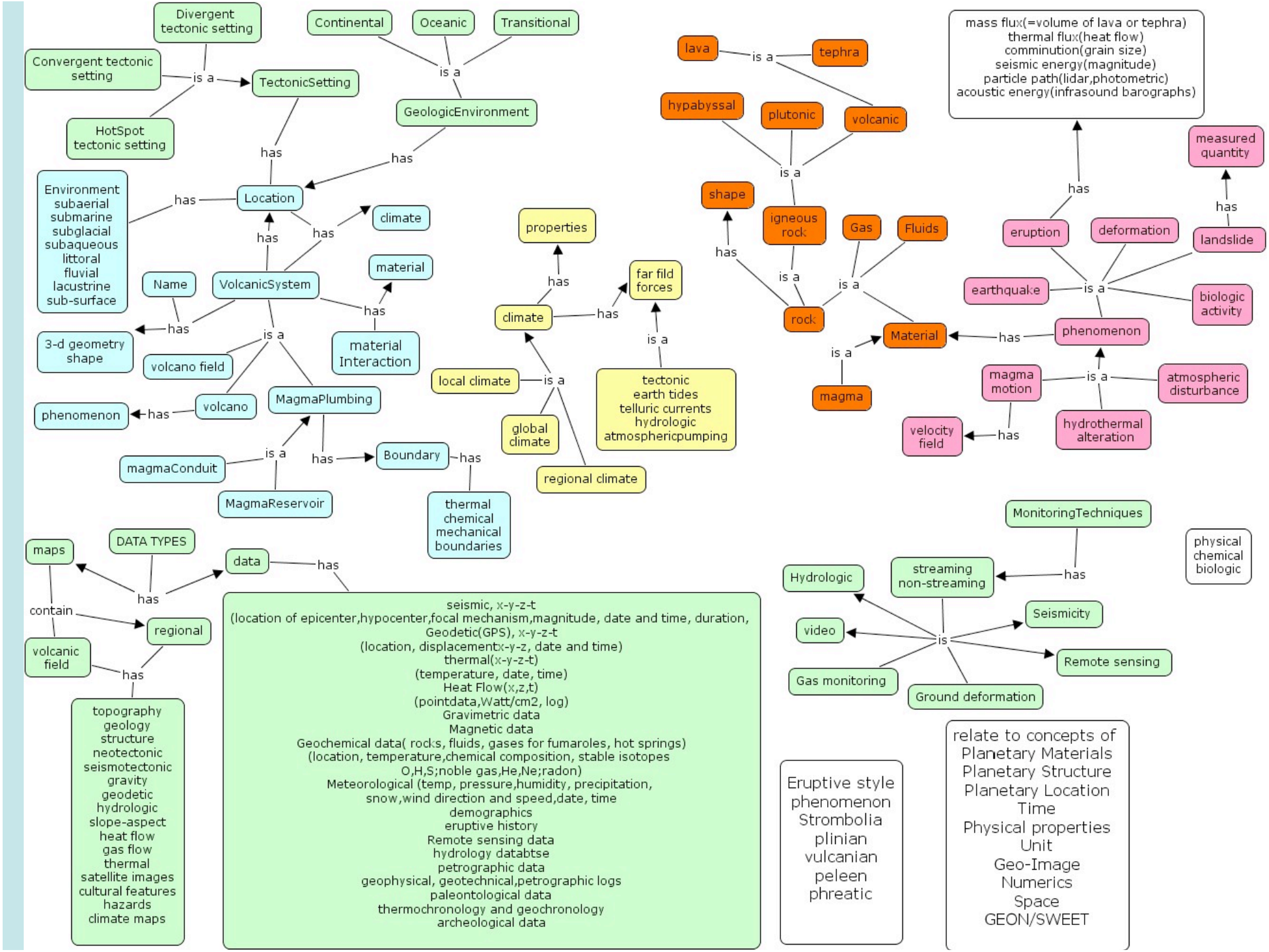
- New release of ODM/MOF
  - Ontology Definition Metamodel/Meta Object Facility (OMG) for UML
  - Provides standardized notation
  - Available for a fee (!) from OMG
  - Books likely to be available soon
- CMAP (concept mapping tool from IHMC)
  - Drag/drop visual development of classes, subclass (is-a) and property relationship
  - Produces an XML file (not OWL)
  - No formal convention
  - Suggest avoid making your CMAP a graph (directed-cyclic)
- White board, text file


a precise point in time
instantaneous events, such as the occurrence of a car accident or the arrival of a package
Instant
InstantEvent
InstantThing
a point or period of time
something that occurs
TemporalEntity
TemporalThing
Event
a period of time bounded by a two instants (that may be the same)
IntervalThing
Interval
IntervalEvent
a period of time bounded by a two instants (that cannot be the same)
events that span some time interval, for example, a meeting.
ProperInterval
ProperInterval Thing


Data analysis methods
Mathematical statistics
Evolutionary modeling
Machine learning
Distribution parameters estimate
Testing of statistical hypotheses
Variance analysis
Correlation analysis
Regression analysis
Time-series analysis
Spectral analysis
Filtration
Correlation analysis
Multivariate analysis
Cluster analysis
Discriminant analysis
Factorial analysis
Scaling
Principal-component method
Genetic algorithms
Artificial neural networks
ART-network
Back propagation network
Hemming network
Hopfield network
Kohonen network
RBF networks
Decision trees
Gain criterion
ID3
C4.5
NewID
Gini
CART


Thing
Individual
PartiallyIntangible
MathematicalOrComputationalThing
PartiallyIntangibleIndividual
Intangible
TemporalThing
SpatialThing
MathematicalThing
IntangibleIndividual
TimeInterval
Situation
MathematicalObject
SetOrCollection
AttributeValue
Set-Mathematical
Collection
SomethingExisting
Situation-Temporal
SpatialThing-Localized
Relation
Event
StaticSituation
PartiallyTangible
PhysicalEvent
TruthFunction
FunctionalRelation
ScopingRelation
TangibleThing
Function-Denotational
FixedArityRelation
Configuration
BinaryRelation
Predicate
IntangibleExistingThing
CompositeTangible AndIntangibleObject
BinaryPredicate
ObjectPredicate
Quantifier
Role
LogicalConnective
BinaryRolePredicate


VSTO ONTOLOGY
LEGEND
Ontology "data" class
Ontology "service" class
Ontology "service impl" class
Ontology "restriction" class
object property
inheritance
inferred inheritance
DateTimeInterval
DataArchive
dataServedBy
servesDataFor +
Instrument
hasInstrumentOperatingMode +
Instrument OperatingMode
NeutralTemperature
Optical Instrument
Interferometer
FabryPerot
hasOperatedInstrument +
isOperatedByObservatory
Observatory
hasMeasuredParameter +
hasMeasuredParameter +
hasCoordinate
Parameter
TimeDependentParameter
hasCoordinate +
Coordinate
Time
Longitude
Altitude
Latitude
TabDataFile
OpendapDataFile
....DataFile
DataFile
TimeSeriesPlot
Attribute
hasAttribute +
DataProduct
DataPlot
GeoPlot
DataImage
BestDataImage
AnyDataImage
AverageDataImage
hasOutput
hasDataProduct +
DataRequest
hasInput
DataConstraint
hasDataRequest
hasConstraint
DataResponse
hasDataResponse
DataService
MetadataService
CEDARdataService
MLSOdataService
CEDARmetadataService
MLSOmetadataService
hasMetadataService
Composite MetadataService


Divergent tectonic setting
Continental
Oceanic
Transitional
Convergent tectonic setting
is a
TectonicSetting
is a
GeologicEnvironment
HotSpot tectonic setting
has
has
Environment
subaerial
submarine
subglacial
subaqueous
littoral
fluvial
lacustrine
sub-surface
has
Location
has
has
climate
has
material
Name
VolcanicSystem
has
has
3-d geometry shape
is a
material Interaction
volcano field
is a
volcano
MagmaPlumbing
phenomenon
has
magmaConduit
is a
has
Boundary
has
MagmaReservoir
thermal
chemical
mechanical
boundaries
properties
has
far fild forces
climate
has
is a
local climate
is a
tectonic
earth tides
telluric currents
hydrologic
atmosphericpumping
global climate
regional climate
lava
is a
tephra
hypabyssal
plutonic
volcanic
is a
shape
igneous rock
Gas
Fluids
has
is a
is a
rock
Material
is a
magma
mass flux(=volume of lava or tephra)
thermal flux(heat flow)
comminution(grain size)
seismic energy(magnitude)
particle path(lidar,photometric)
acoustic energy(infrasound barographs)
has
measured quantity
has
eruption
deformation
landslide
earthquake
is a
biologic activity
has
phenomenon
magma motion
is a
atmospheric disturbance
has
hydrothermal alteration
velocity field
has
maps
DATA TYPES
data
has
has
contain
regional
volcanic field
has
topography
geology
structure
neotectonic
seismotectonic
gravity
geodetic
hydrologic
slope-aspect
heat flow
gas flow
thermal
satellite images
cultural features
hazards
climate maps
seismic, x-y-z-t
(location of epicenter,hypocenter,focal mechanism,magnitude, date and time, duration,
Geodetic(GPS), x-y-z-t
(location, displacementx-y-z, date and time)
thermal(x-y-z-t)
(temperature, date, time)
Heat Flow(x,z,t)
(pointdata,Watt/cm2, log)
Gravimetric data
Magnetic data
Geochemical data(rocks, fluids, gases for fumaroles, hot springs)
(location, temperature,chemical composition, stable isotopes
O,H,S;noble gas,He,Ne;radon)
Meteorological (temp, pressure,humidity, precipitation,
snow,wind direction and speed,date, time
demographics
eruptive history
Remote sensing data
hydrology databtse
petrographic data
geophysical, geotechnical,petrographic logs
paleontological data
thermochronology and geochronology
archeological data
MonitoringTechniques
has
Hydrologic
streaming non-streaming
Seismicity
video
is
Remote sensing
Gas monitoring
Ground deformation
physical
chemical
biologic
Eruptive style
phenomenon
Strombolia
plinian
vulcanian
peleen
phreatic
relate to concepts of
Planetary Materials
Planetary Structure
Planetary Location
Time
Physical properties
Unit
Geo-Image
Numerics
Space
GEON/SWEET

Editors

- Protégé (http://protégé.stanford.edu)
- SWOOP (http://mindswap.org/2004/SWOOP; see also http://www.mindswap.org/downloads/)
- Altova SemanticWorks (http://www.altova.com/download/semanticworks/semantic_web_rdf_owl_editor.html)
- SWeDE (http://owl-eclipse.projects.semwebcentral.org/InstallSwede.html), goes with Eclipse
- Medius

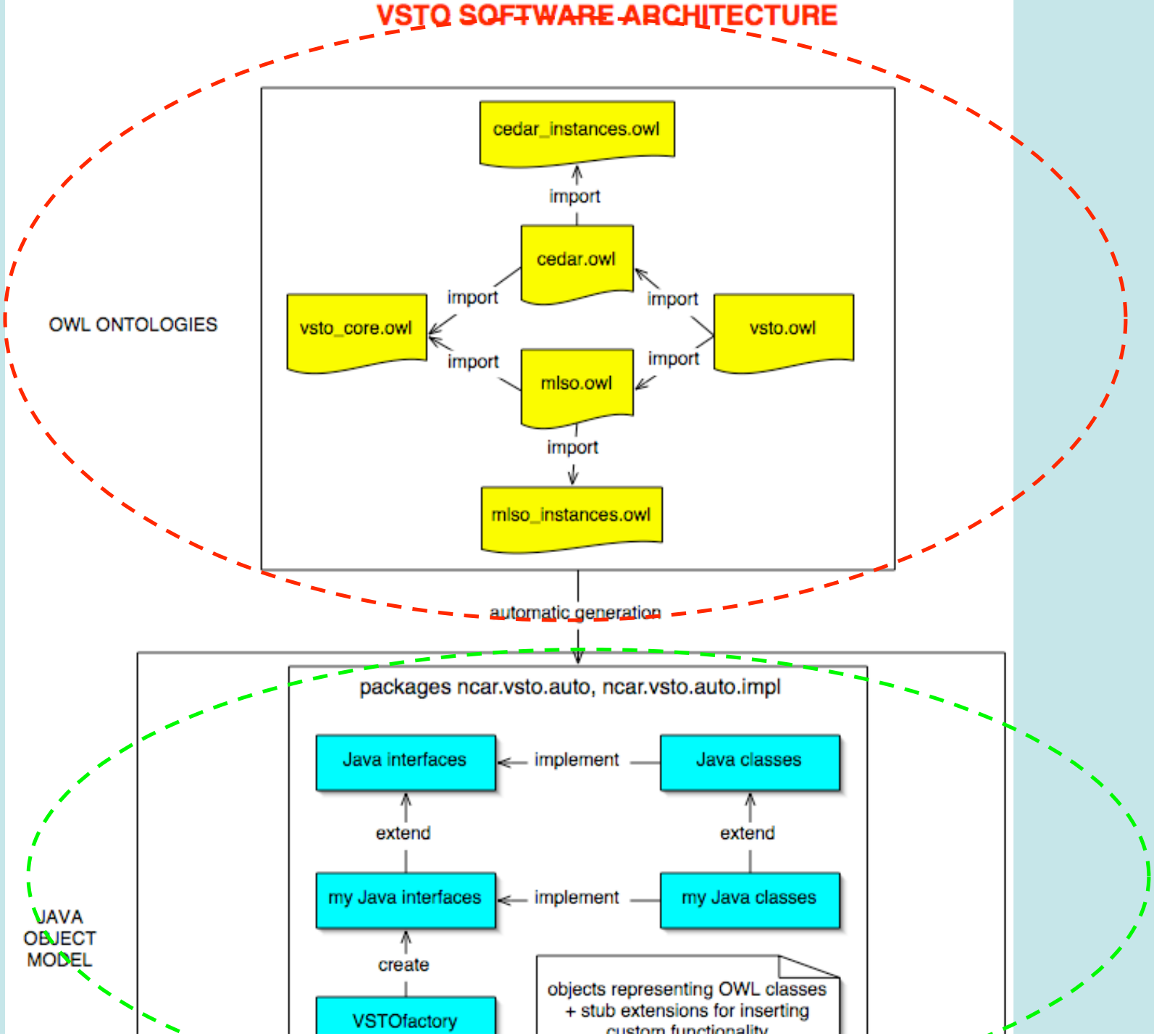
Triple Stores

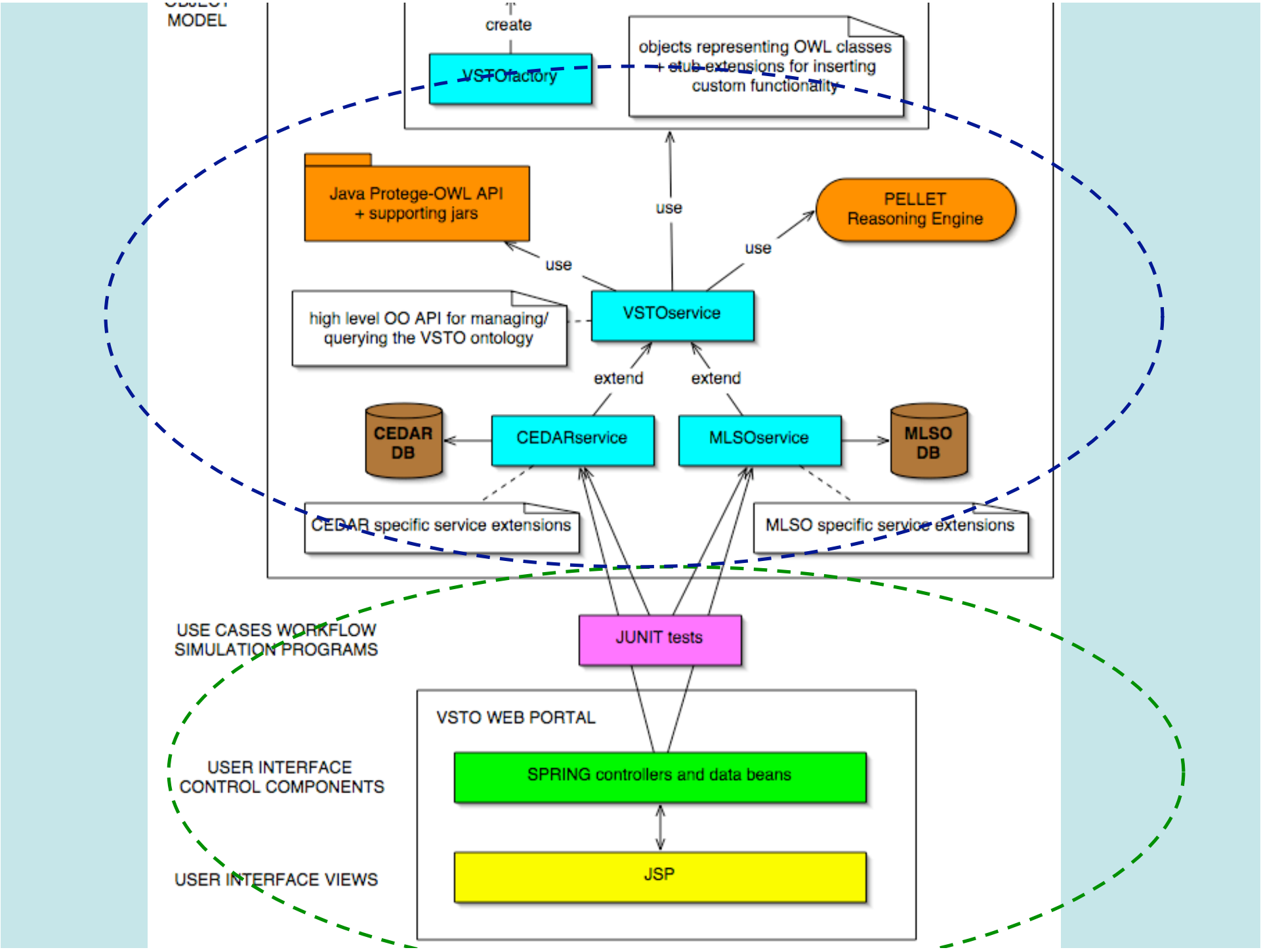
- Jena (http://jena.sourceforge.net/)
- SeSAME/SAIL (http://www.openrdf.org/)
- KOWARI (http://www.kowari.org/) ->
- Mulgara (http://www.mulgara.org/)
- Redland (http://librdf.org/index.html)
- Many others (relational, object-relational)

Software development tools

- Protégé, w/ plug-ins - some better than others
- SWOOP (OWL analyzer, partitioner)
- Jena (http://jena.sourceforge.net/)
- Eclipse (full integrated development environment for Java; http://www.eclipse.org/)

VSTO SOFTWARE ARCHITECTURE




MODEL
create
VSTOfactory
objects representing OWL classes
+ stub extensions for inserting
custom functionality
Java Protege-OWL API
+ supporting jars
use
use
use
PELLET
Reasoning Engine
high level OO API for managing/
querying the VSTO ontology
VSTOservice
extend
extend
CEDAR
DB
CEDARservice
MLSOservice
MLSO
DB
CEDAR specific service extensions
MLSO specific service extensions
USE CASES WORKFLOW
SIMULATION PROGRAMS
JUNIT tests
VSTO WEB PORTAL
USER INTERFACE
CONTROL COMPONENTS
SPRING controllers and data beans
USER INTERFACE VIEWS
JSP

Ph.D. topics ;-)

- Fuzzy logic
  - look at alternatives of Description Logic based on fuzzy logic
  - alternatively, extend RDF(S) with fuzzy notions
- Probabilistic statements
  - have an OWL class membership with a specific probability
  - combine reasoners with Bayesian networks
- Security, trust, provenance
  - combining cryptographic techniques with the RDF model, sign a portion of the graph, etc
- Ontology merging, alignment, term equivalences, versioning, development, ...

Tutorial Summary

- Semantics are in-use in a variety of fields
- Substantial RDF and OWL encodings of knowledge
- Standards are in place in key areas, some not quite
- Tools are in reasonable shape, no killer-tool
- Best practices DO exist, even in Earth Sciences
- PARTNER with someone already familiar
- **A little semantics goes a long way** (DLM)

How to get involved

- ESIP Semantic Web Cluster
  - http://wiki.esipfed.org/index.php/Semantic_Web
  - esip-semanticweb@rtpnet.org, subscribe at http://rtpnet.org/mailman/listinfo/esip-semanticweb
  - Telecon schedule - monthly - 11am (PT). noon (MT), 1pm (CT), 2pm (ET) - 2nd Tuesday starting in January
  - Meeting 8am, Thursday Jan. 4th
  - Semantic Web Technology tracks (Wed/Thu)
  - Joint meetings with Metadata and Web Services
  - Demos and a discussion of SWEET and community ontology development
- NASA/DSWG/TIWG Semantic Web sub-group
  - Semantic Web sub-group meets by telecon, 4th Thursday: 4pm ET/1pm PT as part of TIWG
  - http://teambps.mywsssite.com/seeds/wg/infusion/default.aspx

NASA/DSWG/TIWG

- Semantic Web Technology Infusion Roadmap
- Version 0.5
- Soliciting feedback

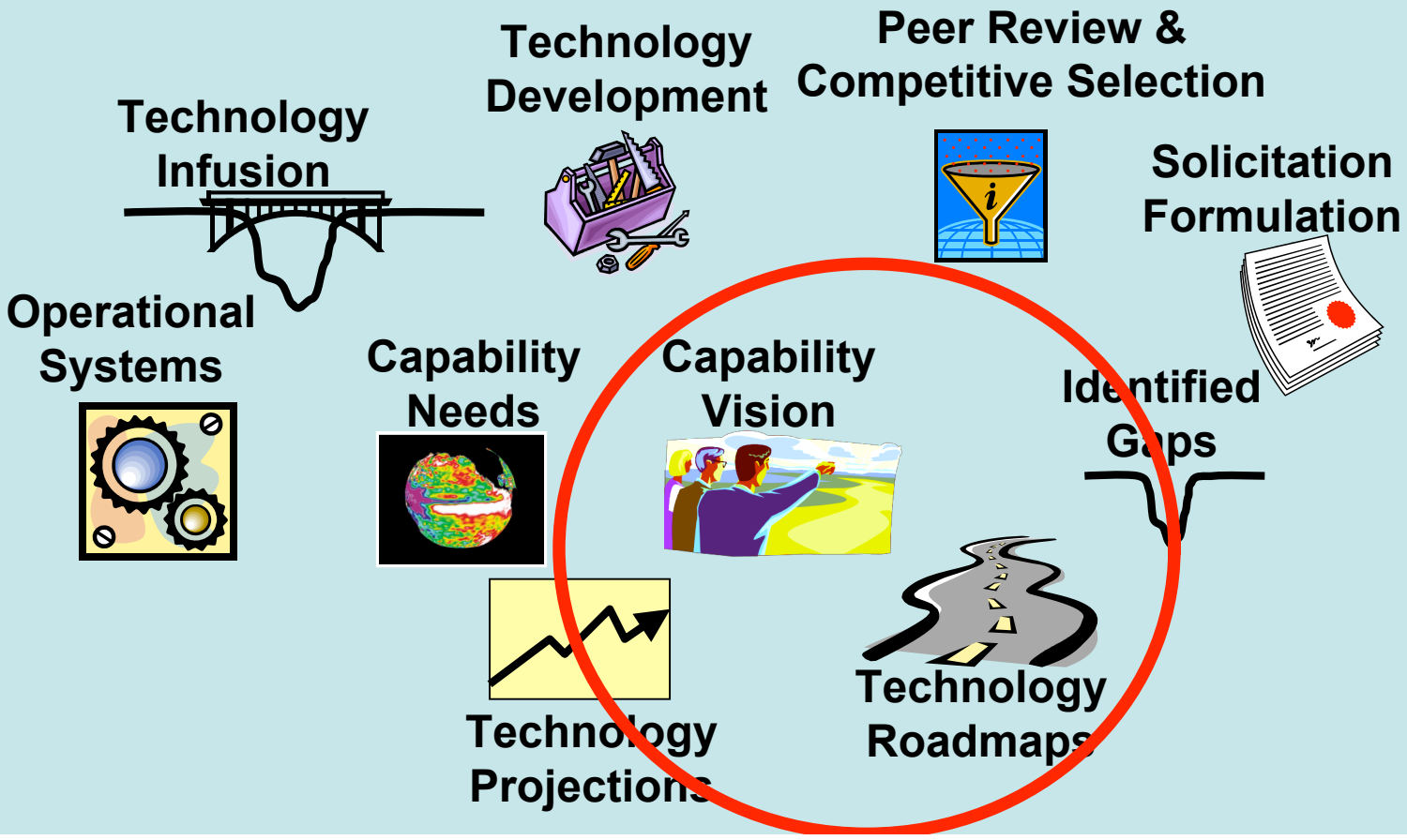
Terminology

- Ontology (n.d.). The Free On-line Dictionary of Computing. http://dictionary.reference.com/browse/ontology
  - An explicit formal specification of how to represent the objects, concepts and other entities that are assumed to exist in some area of interest and the relationships that hold among them.
- Semantic Web
  - An extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation, www.semanticweb.org
  - Primer: http://www.ics.forth.gr/isl/swprimer/
- Semantic Grid
  - Semantic services to use the resources of many computers connected by a network to solve large scale computational problems
- Technology Infusion Roadmap
  - Timeline for widespread adoption of key technologies including tailoring to Earth science needs
- *Service Oriented Architecture, Web Services, Service Chaining - see Web Services Roadmap*

- Languages
  - OWL - Web Ontology Language (W3C)
  - RDF - Resource Description Framework (W3C)
  - OWL-S/SWSL - Web Services (W3C)
  - WSMO/WSML - Web Services (EC/W3C)
  - SWRL - Semantic Web Rule Language
  - PML - Proof Markup Language
  - ODM/MOF - Ontology Definition Metamodel/Meta Object Facility (OMG)
  - Editors: Protégé, SWOOP, Medius, SWeDE
- Reasoners
  - Pellet, Racer, Medius KBS, FACT++, fuzzyDL, KAON2, MSPASS, QuOnto
- Query Languages
  - SPARQL, XQUERY, SeRQL, OWL-QL, RDFQuery
- Other Tools for Semantic Web
  - Search: SWOOGLE swoogle.umbc.edu
  - Collaboration: www.planetont.org
  - Other: Jena, SeSAME/SAIL, Mulgara, Eclipse, KOWARI
  - Semantic wiki: OntoWiki, SemanticMediaWiki
- Emerging Semantic Standards for Earth Science
  - SWEET, VSTO, MMI, GeoSciML

Background: Technology Infusion Process

- Established a capability vision for Earth science information systems
- Identified Interoperable Information Services as a key capability in the vision
- Identified semantic web as one of the primary supporting technologies
- Currently defining a roadmap for semantic web technology infusion



Infusing Semantic Web Will Help Realize the Vision for all Middleware Services and Assisted Capabilities

Scalable Analysis Portals	Community Modeling Frameworks	Assisted Data & Service Discovery	Assisted Knowledge Building

Interactive Data Analysis	Seamless Data Access	Interoperable Information Services	Responsive Information Delivery	Verifiable Information Quality

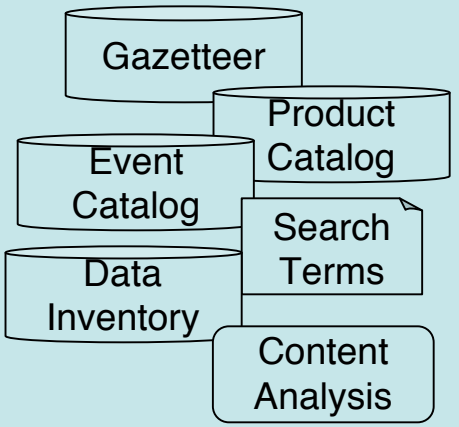
Evolvable Technical Infrastructure

Semantic Web Technology Infusion
(OWL, OWL-S, SWSL, SWSO)
(SWSF, Jena, Eclipse, SeSAME, etc.)

Assisted Data & Service Discovery

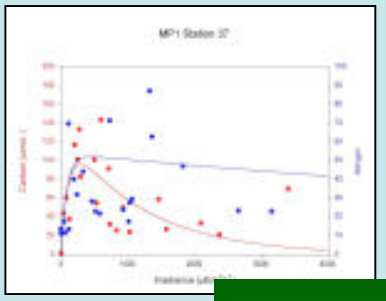
- Vision
  - Semantic and cont
    researchers to quic
    information needed
    analysis using prop
    specific jargon, and
    specifications
- *Annotation - detail of how t ge bw vision and enabling tech….*
- *Enabling technologies*
  - *Data and service description standards (XML, WSDL, RDF, OWL,OWL-S, DAML), web service directories (UDDI), syndication services (RSS), topic maps*
  - *Rule-based logic systems*
  - *Established directory services (GCMD, ECHO, THREDDS)*

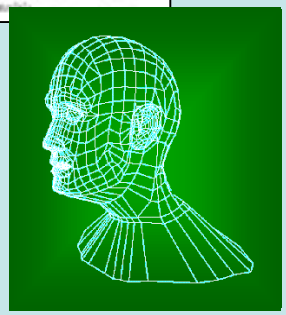
Value added nature of semantics, reasoning, query expansion, term narrowing, unifying/abstracting selection workflows, concepts, ontologies, domain specific - enable…. SWEET, others.



Assisted Knowledge Building

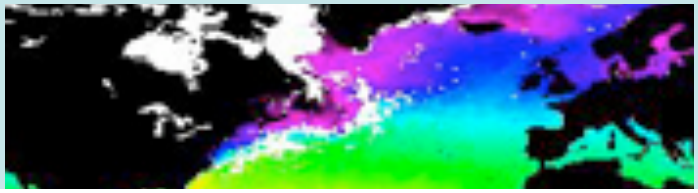
- Vision
  - Provide research and operations assistance using intelligent systems
- *Enabling technologies*
  - *Data mining algorithms (Support vector machines, independent component analysis, rule induction)*
  - *Data mining toolkits (Adam, D2K, Darwin)*
  - *Data mining plug-ins (IMAGINE, ENVI, ArcGIS)*
  - *Data and service description standards (XML, WSDL, RDF, OWL, DAML), web service directories (UDDI), syndication services (RSS), topic maps*
  - *Rule-based logic systems*



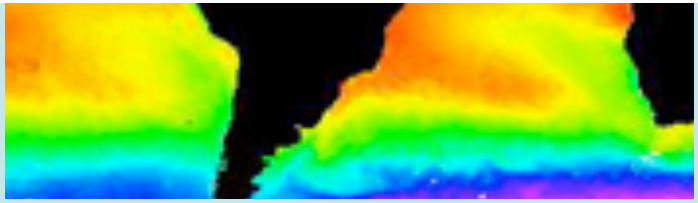


Interactive Data Analysis

- Vision
  - Reduce research algorithm implementation from months to hours
- *Enabling technologies*
  - *Visual grammars*
  - *Visual programming environments (Cantata, Triana, Grist/Viper, Wit)*
  - *High-level analysis tools (IDL, Matlab, Mathematica)*

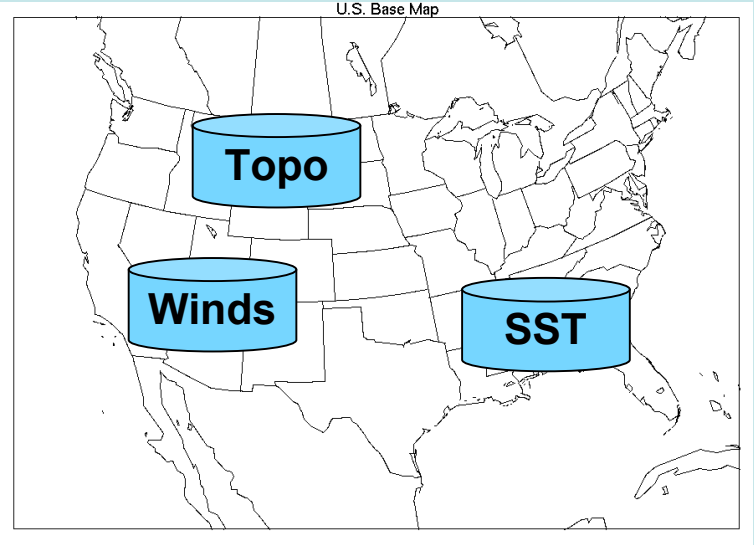


$$\rho C_{Pg} u \frac{\partial T}{\partial x} = \lambda \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} \right) + G$$



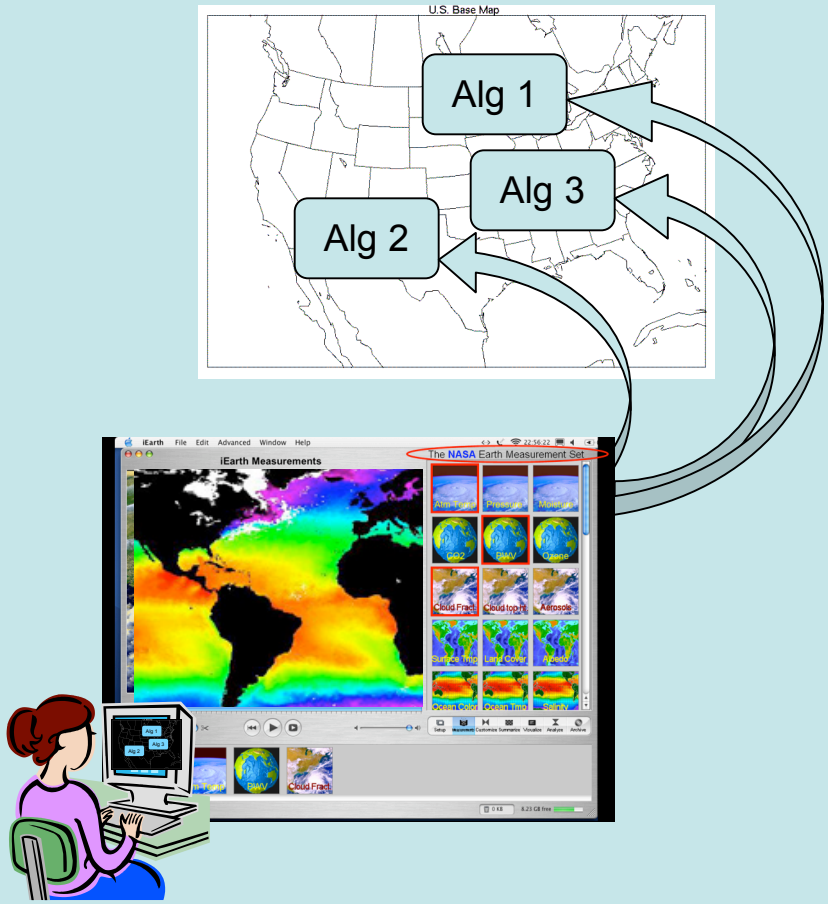
Seamless Data Access

- Vision
  - Users can access current data from authoritative sources from any programming environment or analysis tool regardless of the data's physical location
- *Enabling technologies*
  - *Network data access protocols (OpenDAP, WMS/WCS, WebDAV, GridFTP)*
  - *Established data server tools (MapServer, DODS/LAS, ArcWeb)*
  - *Semantic metadata (OWL-S)*



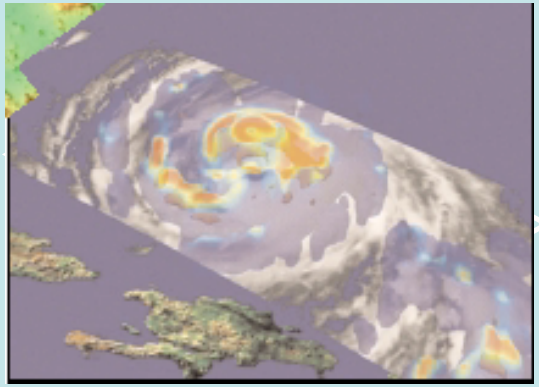
Interoperable Information Services

- Vision
  - Increase synergy in the Earth science community by leveraging in-place resources and expertise to provide information services on demand
  - Allows researcher to simply invoke remote services from within a local analysis tool
- *Enabling technologies*
  - *Network service protocols (SOAP, Java RMI, OPeNDAP, CORBA)*
  - *Utility/grid computing protocols & toolkits (Globus)*
  - *Semantic metadata (OWL-S)*



Responsive Information Delivery

- Vision
  - Ensure research priorities are met and enable new uses of Earth science data
- *Enabling technologies*
  - *Optical networks (National LambdaRail)*
  - *Peer-to-peer networks with swarming (Modster)*
  - *Direct downlink (MODIS/AIRS DDL)*

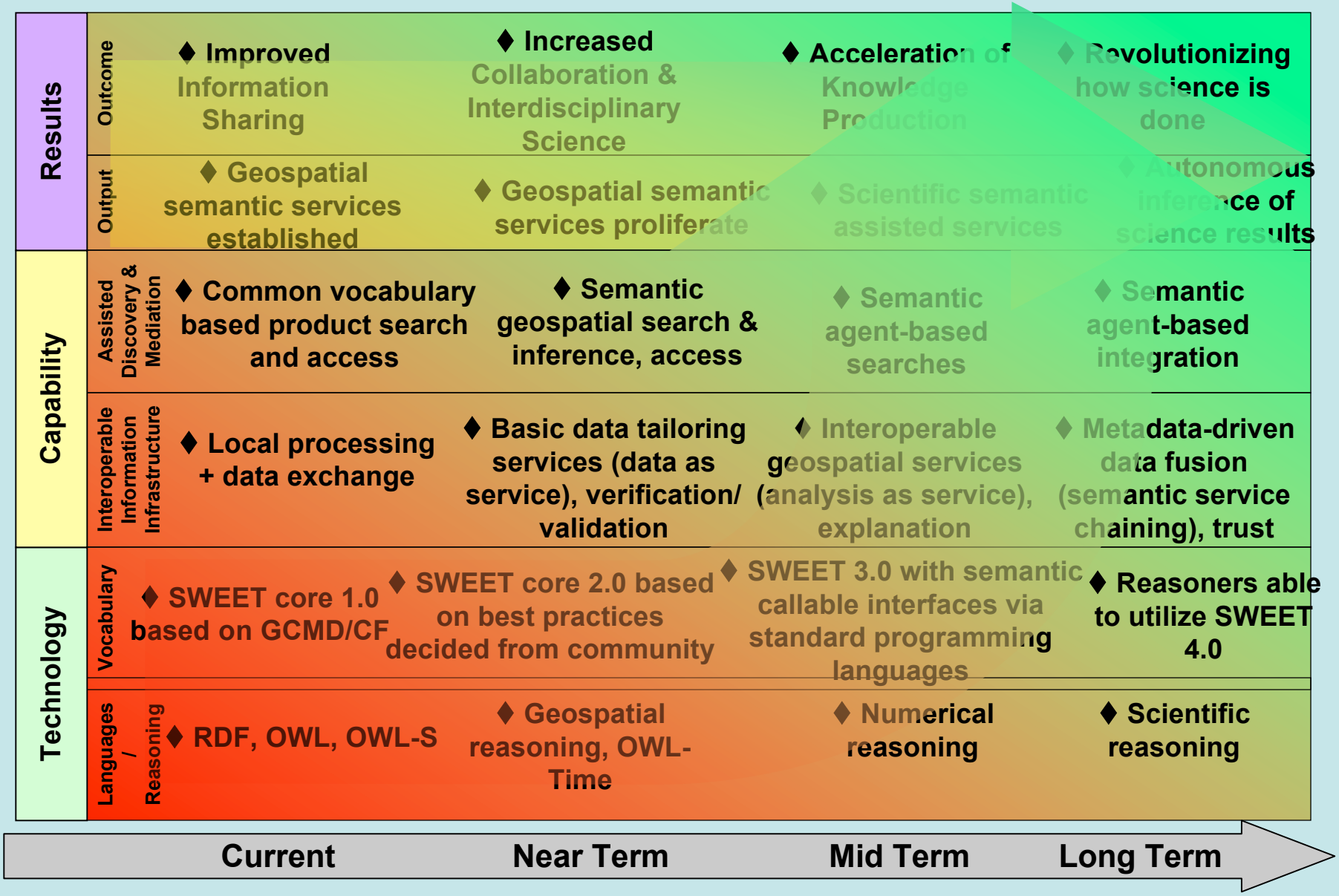


Verifiable Information Quality

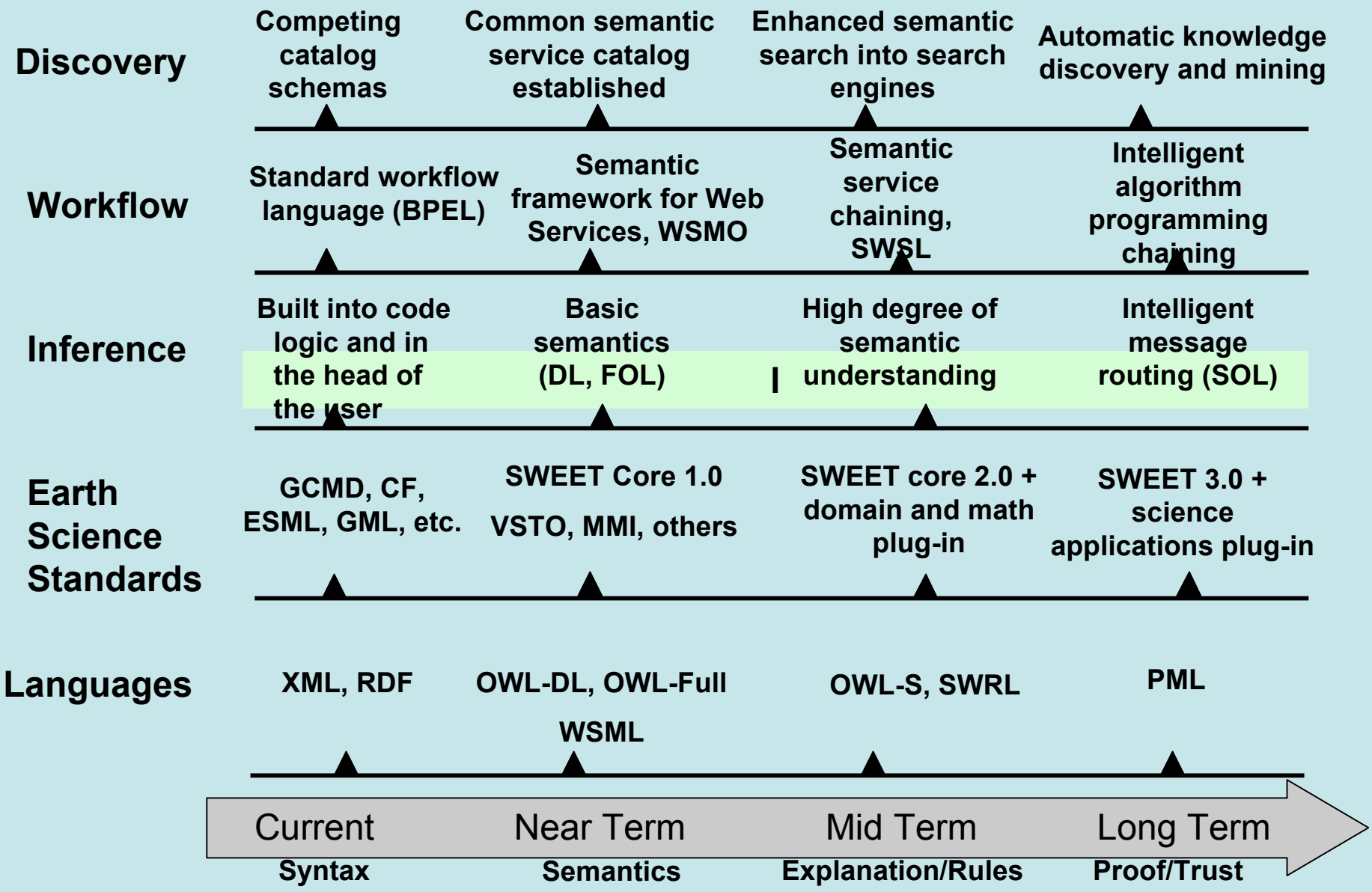
- Vision
  - Provide co
    informatio
    and enabl
    communit
    provider r
- *Enabling te*
  - *Data pedi*
    algorithm
  - *Machine-r*
    formats (XML)
    semantics (OWL-

Value added nature of semantics
Need to be able to describe the various dimensions information quality in a consistent way.
Some work in progress on ontologies for information quality
http://www.qurator.org/ publications
Managing Information Quality in e-science
http://euindia.dimi.uniud.it/wo6presentation/Missier.ppt

Semantic Web Roadmap



Semantic Web: Roadmap Details



Ongoing Steps

- Refine roadmap milestones
  - Identify issues specific to Earth science
  - Review dependencies and revise milestone order
- Formulate recommendations
  - Recommend actions that NASA can take to help achieve milestones (to ESTO by April, 2007)
- Start the infusion process
  - Identify technology candidates that could be submitted to the DSWG standards process to accelerate infusion
  - Identify technology champions (groups and individuals) to shepherd technologies through the process, e.g. ESIP semantic web cluster